ASSESSING THE BIOLOGICAL QUALITY OF FRESH WATERS:

RIVPACS AND OTHER TECHNIQUES

Edited by

JOHN F. WRIGHT
Institute of Freshwater Ecology, UK

DAVID W. SUTCLIFFE
Freshwater Biological Association, UK

MIKE T. FURSE
Institute of Freshwater Ecology, UK

Invited contributions from an International Workshop held in Oxford, UK

on 16-18 September 1997 by the

Institute of Freshwater Ecology (NERC Centre for Ecology and Hydrology), UK

Environment Agency, UK

Environment Australia

Land and Water Resources R&D Corporation, Canberra, Australia

Published by the
FRESHWATER BIOLOGICAL ASSOCIATION,
AMBLESIDE, CUMBRIA, UK

ISBN 0-900386-62-2

FBA Special Publications:

A natural history of the lakes, tarns and streams of the English Lake District
by G. Fryer

Eutrophication:
Research and application to water supply
edited by D. W. Sutcliffe & J. G. Jones

Water quality and stress indicators in marine and freshwater ecosystems:
Linking levels of organisation (individuals, populations, communities)
edited by D. W. Sutcliffe

The microbiological quality of water
edited by D. W. Sutcliffe

Some English lakes as diverse and active ecosystems:
A factual summary and source book
compiled and edited by J. F. Talling

CHAPTER 1

An introduction to RIVPACS

JOHN F. WRIGHT

Institute of Freshwater Ecology, River Laboratory, East Stoke, Wareham, Dorset, BH20 6BB, UK

Summary

RIVPACS (River InVertebrate Prediction And Classification System) is a software package developed by the Institute of Freshwater Ecology (IFE). The primary application is to assess the biological quality of rivers within the UK. RIVPACS offers site-specific predictions of the macroinvertebrate fauna to be expected in the absence of major environmental stress. The expected fauna is derived by RIVPACS using a small suite of environmental characteristics. The biological evaluation is then obtained by comparing the fauna observed at the site with the expected fauna.

RIVPACS also includes a site classification based on the macroinvertebrate fauna of the component reference sites. New sites, judged by their fauna to be of high biological quality, may be allocated to classification groups within the fixed RIVPACS classification. This has potential for evaluating sites for conservation.

In this chapter, the origins and history of the RIVPACS approach are described, including major scientific and operational developments over the life of the project. RIVPACS III is described in detail and predictions at different taxonomic levels are demonstrated. The value of the reference dataset for river management and conservation is examined, and the chapter concludes with a brief consideration of some future challenges.

Introduction

Background

The initial four-year project which eventually lead to the development of RIVPACS started in October 1977, funded by the Natural Environment Research Council and the Department of the Environment. The project had two major objectives: (1) development of a biological classification of unpolluted running-water sites in Great Britain, based on the macroinvertebrate fauna, and (2) assessment of whether the macroinvertebrate community at a site could be predicted using physical and chemical features.

Careful decisions were made on the choice of reference sites, the field and laboratory procedures and the methods for analysis (Wright, Moss *et al.* 1984). Decisions made over 20 years ago on the field-sampling protocol and the level of identification used at reference sites have determined the present field-sampling protocol (Murray-Bligh, Furse *et al.* 1997) and resulted in a wide range of prediction options in RIVPACS III (Wright, Moss *et al.* 1997).

However, the long-term success of the project depended upon the initial demonstration of a strong relationship between the environmental features and faunal characteristics of the original 268 reference sites (Wright, Moss *et al.* 1984). This was essential for an operational system that would be capable of delivering reliable predictions of the fauna over a wide range of river types. During RIVPACS development, practical issues of concern to end-users were addressed, to ensure that biological monitoring gained a higher profile alongside chemical monitoring. A number of these issues, including quality assurance, presentation of biological results, and methods for detecting spatial and temporal change, are described in later chapters of this volume.

The accumulated experience of the IFE team, together with the new ideas being explored by several groups in other countries (e.g. Reynoldson, Bailey *et al.* 1995; Norris 1996), make this the ideal time to assess current achievements and examine some future challenges.

The context for RIVPACS

Biological monitoring is required, in addition to chemical monitoring, because the ultimate consequences of environmental stress can only be determined by an appraisal of the biota. The 16th Report of the Royal Commission on Environmental Pollution (1992) dealt with Freshwater Quality and their first recommendation was that *"The regulatory authorities should endeavour to develop a general classification scheme based on biological assessment for use throughout the UK in the 1995 and subsequent river quality surveys"*. The question of which major taxonomic group to use for site appraisal has been addressed on many occasions (Hellawell 1986). Although it would be wise to retain an open mind when designing a monitoring programme for a new region (Ormerod, Rundle *et al.* 1994), the advantages of using macroinvertebrates have made them the favoured group in most surveillance and monitoring studies (Rosenberg & Resh 1993b). Wallace & Webster (1996) review the essential role played by macroinvertebrates in the functioning of stream ecosystems.

The effects of stress in running waters can be detected at many different levels of organisation, ranging from biochemical and physiological effects on individuals, through the response of a population, to community responses and ecosystem effects (Sutcliffe 1994). A comprehensive monitoring programme may include a number of these elements. Bunn (1995) pointed out that the ecological consequences of stress are best examined at higher levels of organisation, but these assessments detect impacts after the event. A future challenge will be the integration of the community approach with early warning techniques based on lower levels of organisation. In the meantime, biological surveillance of communities, with special emphasis on characterising taxonomic richness and composition, is perhaps the most sensitive tool now available for quickly and accurately detecting alterations in aquatic ecosystems (Cairns & Pratt 1993).

It would be invaluable to understand the processes that result in the observed patterns of community structure in unstressed river systems. This would provide a firm basis for interpreting the mechanisms leading to changes in the structure and functioning of communities under stress. At present, this knowledge is very restricted (Hildrew 1992) because a wide variety of ecological and evolutionary processes, historical events and geographical circumstances, contribute to the patterns of species richness and composition observed in nature (Cornell & Lawton 1992; Schluter & Ricklefs 1993). As a consequence, lotic community responses to a range of environmental stresses are well documented, but the processes of community change are poorly understood.

Hellawell (1986) proposes three major categories of environmental stress. These are natural stresses (e.g. droughts and floods), imposed stresses (e.g. sewage pollution, toxic waste and pesticides) and environmental manipulation by man (e.g. reservoir construction, channel modification and the transfer of water between catchments). The macroinvertebrate fauna may be affected by each one of these stresses, and the fauna at any given site may be the end result of more than one category of stress.

A review of the longer-term trends expected in stressed ecosystems is given by Odum (1985). These include a consideration of energetics, nutrient cycling, community structure and general system-level trends. At the community level, Odum suggests that the proportion of r-strategists will increase, the size of organisms will decrease, lifespans will decrease, food chains will shorten and, typically, species richness will decrease, allowing some taxa to dominate. He hypothesises that functional properties such as community metabolism are more robust than species composition and other structural properties. Perry (1994) also discusses the idea that functional attributes such as production, respiration and nutrient cycling may be more robust than structural attributes, including species composition and richness. If this is true, then monitoring structural changes may give earlier warning of the effects of environmental stress than the measurement of functional attributes.

In RIVPACS, the emphasis is on the prediction of taxonomic composition and richness. In RIVPACS III, attempts have also been made to predict $\log_{10}$ categories of abundance at family level in order to detect early signs of structural change before substantial loss of taxa.

In order to develop a general system in which the fauna may be used to detect environmental stress, three features must be present. First, a mechanism is required for predicting the "expected" fauna at a given unstressed site, in order to provide a "target" community. Second, the procedure must be applicable to a wide range of sites. Finally, there must be an appropriate procedure for comparing the observed fauna with the expected fauna.

The RIVPACS approach

The basic concept and use of reference sites

RIVPACS offers a prediction of the "expected" fauna at a given site, with stated environmental features, through a procedure which draws on information from a series of appropriate sites of high biological quality. In practice, no rivers in Great Britain are unaffected by human activity in the catchment or in the river channel itself. Thus, there is a need to identify those rivers and sites that are minimally impacted and the best examples of their type. At the outset, detailed discussions took place between biologists throughout Great Britain and the IFE team, in order to generate a list of potentially suitable river systems. Before sampling began, a river and site-selection procedure was devised, based on geological, physical and chemical factors to ensure a wide coverage of different river types (Wright, Moss *et al.* 1984; Chapter 20).

The funding arrangements following the initial study necessitated a step-by-step approach to building a comprehensive series of minimally stressed reference sites. Increasing experience of the fauna to be expected at high quality sites, coupled with more severe criteria for site acceptance (Wright, Furse *et al.* 1995), resulted in a comprehensive set of "reference sites" (Hughes 1995; Reynoldson, Norris *et al.* 1997) with high capability for setting a target of the fauna to be expected at a given site. The benefits and difficulties associated with the acquisition and use of reference sites are discussed by Reynoldson & Wright in Chapter 20. Prior to each major analysis, macroinvertebrate and environmental data were collated for each reference site. An essential feature of the RIVPACS approach is the classification of reference sites using the macroinvertebrate fauna. Thus, sites with similar macroinvertebrate assemblages are brought together and no assumptions are made about the environmental features that influence species occurrence. This approach differs from the strategy adopted in the Rapid Biological Assessment method used in the USA, in which environmental attributes that indicate "relatively undisturbed" sites are the basis for selecting a series of reference sites in a given ecoregion or sub-ecoregion (Omernik 1995). For this, it is assumed that reference sites in the defined region have relatively similar macroinvertebrate assemblages which may be used to calculate an array of measures or metrics that define the expected condition, and against which the biological condition of the test sites may be determined (Barbour *et al.* 1995; Chapter 19).

Predicting the expected fauna

The next stage in RIVPACS was to determine whether classification groups defined on biological criteria were also coherent with respect to a series of environmental attributes. This was investigated using multiple discriminant analysis (MDA) (Klecka 1975). Ideally, the environmental attributes should be relatively stable over time and unaffected by environmental stress. A description of this procedure, and details of the effectiveness of the technique for demonstrating the relationship between the biological and environmental features of the reference dataset, are given by Moss in Chapter 2.

The detailed techniques whereby environmental features for a new site were used in RIVPACS to predict the expected fauna of that site, are also presented in Chapter 2. Essentially, this was a two-stage operation. First, the probability of the new site being in each one of the classification groups was determined. In practice, a site would normally have a high probability of occurrence in one or a few classification groups and all others would not make a significant contribution. Second, a novel technique was used to predict the probability of occurrence of individual taxa at the site (Moss, Furse *et al* 1987). This required information on the frequency of occurrence of taxa in each classification group and the probabilities of classification group membership for the new site.

Comparing the observed and expected fauna at each site

RIVPACS offers a comparison between the fauna captured at the site using the standard field protocol and the expected fauna derived by prediction. The observed fauna is represented by positive occurrences of taxa, whereas the expected fauna is displayed as probabilities of capture ranging from 100% probability to the percentage probability requested by the user. The latter is normally 0.1% for family-level predictions, but frequently 50% for species-level predictions. At family level, the ratio of the observed to the expected number of taxa is easily computed, and provides the first indication of whether the fauna matches expectation. Examples of observed/expected (O/E) ratios are given later.

Some practical considerations

Sampling methods

Perhaps the most crucial decisions made at the outset of the project were how to sample, where to sample, and when to sample. Early consultations with Water Authority and River Purification Board biologists indicated that pondnet sampling was widespread but that local procedures varied in detail. In some regions, sampling was confined to riffles because the detection of organic pollution was a priority and riffle assemblages were known to be sensitive to this form of stress (Balloch *et al.* 1976). In other regions, a variety of habitats were included in the sampling procedure. The need for a simple and flexible sampling procedure for use at a wide range of locations, a classification of reference sites based on comprehensive faunal assemblages, and a prediction system capable of detecting a wide range of environmental stresses, indicated that a single habitat protocol would be inappropriate. On the other hand, separate sampling units for each major habitat at a given site would increase sampling effort and result in non-uniform effort across the sites. It was concluded that a reasonably comprehensive species list for each site would be obtained from a timed pondnet sample in which all major habitats were sampled, approximately in proportion to their occurrence. The procedure for each reference site involved a 3-minute pondnet sample plus a 1-minute manual search in each of spring (March–May), summer (June–August) and autumn (September–November), in order to capture the major components of the fauna.

Using this procedure, in a pilot study, three biologists each took two samples at four sites on one river (Furse, Wright *et al.* 1981). Clustering and ordination of species-level data gave strong site-faithfulness for the six samples from each site, despite the fact that there were significant inter-operator differences in the number of taxa captured ($p < 0.05$) and the study was confined to one season.

At a small number of deep lowland sites, sampled later in the project, a standard pondnet (900 µm mesh, 230x255 mm frame, 275 mm bag depth) on a 1.5 m handle was ineffective. In these cases, a light-weight version of the Medium Naturalist's dredge (Holme & McIntyre 1971; Furse, Moss *et al.* 1986) was used in the deep water in conjunction with pondnetting of marginal areas. Further details of all field procedures may be found in Murray-Bligh, Furse *et al.* (1997).

Reference sites and identification protocols

In the initial programme, sampling sites were chosen at *ca* 5, 10, 20, 30 and 40 km from source and thereafter at 20 km intervals downstream, because rate of change in community composition is greater near the source of each stream (Verneaux 1976). The precise location of each sampling site was always chosen in consultation with the local biologist. Later in the project, further advice was taken on the geographical areas and river types which required greater representation and on the availability of both deep lowland and small stream sites, to make the system more comprehensive and reliable.

The involvement of regional biologists in selecting reference sites and collecting field samples was crucial in building each new version of RIVPACS. In addition, it was essential that samples from reference sites were subjected to a standard laboratory protocol and identified to the same taxonomic level. For this reason, all samples were preserved and sent to the River Laboratory where the IFE team was responsible for processing the samples (Furse, Wright *et al.* 1981) and identifying the fauna to a specified taxonomic level. Where closely related species or genera could not be distinguished consistently, they were represented as a taxonomic "group". Prior to the analyses, the combined seasons taxon lists for each reference site were subjected to a "standard edits" program which standardised the precision of the faunal listings, thus ensuring valid comparison between sites (Wright, Blackburn *et al.* 1996). For each reference site in each season, $\log_{10}$ categories of abundance at family level (i.e. <10, <100, <1000 and 10,000 individuals per family) were recorded to supplement the presence/absence data in the standardised taxon lists.

Classification of reference sites

The procedure chosen for classifying the reference sites by their macroinvertebrate assemblages was two-way indicator species analysis (TWINSPAN) (Hill 1979a), a divisive polythetic technique. Gauch & Whittaker (1981) compared a number of hierarchical procedures (agglomerative and divisive) and took the view that TWINSPAN was an appropriate technique for complex, "noisy", large or unfamiliar datasets. Initially, the standardised taxon lists based on three seasons combined sampling were used to develop site classifications because they demonstrated higher predictive ability than alternative classifications. The latter included the standardised lists for single seasons (qualitative data only), combined/single season classifications at family level (qualitative or $\log_{10}$ categories of abundance) and at BMWP family level (qualitative only) (Furse, Moss *et al.* 1984). Further information on the Biological Monitoring Working Party (BMWP) score system for the appraisal of running-water sites (National Water Council 1981) will be presented later.

Recently, a number of other classification techniques have been used to provide a suitable framework for developing prediction systems (Chapter 2; Moss, Wright *et al.* 1999; Norris 1996). Despite this, TWINSPAN remained the method of choice in RIVPACS III because alternative procedures failed to divide this substantial dataset into classification groups that were coherent with respect to the environmental attributes chosen for prediction (Wright, Furse *et al.* 1995; Wright, Moss *et al.* 1997).

The classification forms a framework for the prediction system, but also may be used to classify new sites based on taxon lists obtained in spring, summer and autumn, using the standard protocol. TWINSPAN offers a "key" to the classification based on a limited number of "differential" taxa that are diagnostic for each division in the classification (Hill 1979a). This key was used to classify new sites in RIVPACS I. However, in RIVPACS II and RIVPACS III, an improved procedure has been adopted (Rushton 1987) in which the full taxon list for a new site is used to generate probabilities of classification group membership. This procedure may be used with standardised taxon lists or BMWP family-level data.

Environmental variables for prediction

Initially, a large number of environmental attributes were acquired for each site. Whereas some variables were time invariant (e.g. altitude, slope, distance from source), others varied with the seasons (e.g. river width, depth, substratum) and for the latter, mean values were derived from observations taken in three seasons. In contrast, chemical variables were represented as annual mean values derived from the best available data.

Although 28 environmental variables were used in the first analyses that linked environmental features to site groups based on faunal characteristics (Wright, Moss *et al.* 1984), it was apparent that many of the variables were highly correlated and therefore redundant. Moss, Furse *et al.* (1987) used a modified list of 28 variables but demonstrated that a subset of these was capable of delivering a practical system. In consequence, changes were made to successive versions of RIVPACS in the search for a limited suite of environmental variables that could deliver a reliable prediction system. The importance of quantifying errors in the measurement of environmental variables is emphasised in Chapter 3 and some thoughts on alternative variables for prediction are considered at the end of this chapter.

Stages in the development of RIVPACS

Initial stages

The major developments and outputs from each phase of the RIVPACS project are presented in Table 1.1. Between 1977 and 1981, the initial field sampling programme was designed and undertaken in collaboration with colleagues in the water industry. IFE staff then identified the macroinvertebrate fauna and collated the environmental data for 268 reference sites, before analyses commenced. A TWINSPAN classification of sites, using standardised species-level data for three seasons, was developed to 16 groups. MDA was then used to predict the group membership of the 268 sites, based on 28 environmental variables. In an internal (re-substitution) test, 76.1% of sites were classified correctly (Wright, Moss *et al.* 1984). Fifteen additional classifications were also constructed, using data from single/combined seasons and different taxonomic levels. These included qualitative data at species, family and BMWP family level, and also family $\log_{10}$ category data. In each case, predictive ability based on the same 28 variables was lower than the species-level combined seasons classification (Furse, Moss *et al.* 1984).

Table 1.1. *Historical review of RIVPACS development, 1977-1997. See the text for further details.*

Dates	No. sites	Main developments	Major outputs
1977-1981	268	Formulation of standard field procedures. Production of a site classification based on the fauna. Demonstration of the strong link between biological and environmental features using MDA.	
1981-1984	370	Increase in geographical coverage of reference sites. Development of a new procedure for predicting the probability of taxon occurrence.	
1984-1988	370	Incorporation of biological indices from the BMWP system into the prediction system.	RIVPACS I tested by water industry biologists.
	438	Increase in the number of deep lowland rivers and small streams. Development of a new classification/prediction system using 438 sites in 25 classification groups (basis for RIVPACS 11).	
1989-1990	438	Operational development of RIVPACS II.	RIVPACS II used in 1990 River Quality Survey.
1991-1995	438	Development of a banding system to summarise results of 1990 River Quality Survey. Comprehensive testing of RIVPACS II using an independent dataset of high quality sites.	
	614+70	Increase in reference sites to give comprehensive coverage in Great Britain, plus a new senes of 70 reference sites for Northern Ireland. Investigation of alternative procedures for site classification and prediction of the fauna, using environmental attributes.	
		Development of a new classification and prediction system for Great Britain and Northern Ireland, based on the enlarged dataset (RIVPACS III).	RIVPACS III used in 1995 General Quality Assessment Survey.
1995-1997	614+70	Development of new procedures for detecting statistically significant temporal and spatial changes.	RIVPACS III+.

The successful demonstration of a strong link between faunal composition and the environmental features of the 268 reference sites led to further funding. The objectives for the next phase (1981–1984) included greater representation of rivers in Wales and Scotland, and lowland rivers in England, reducing the number of variables for prediction and refining the prediction procedure. At first, the environmental variables were used to predict probabilities of classification group membership only. In this phase, a novel technique was devised for predicting the probabilities of capture of individual taxa, based on their frequency of occurrence in the relevant classification groups (Moss, Furse *et al.* 1987). The same paper demonstrated that reducing the number of predictive variables from 28 to 11 (or even 5) resulted in only a modest loss of accuracy.

RIVPACS I

By 1984, a 370-site classification to 30 groups and a revised prediction system were in place. As before, the classification used species-level data and predictions were made at this taxonomic level. However, in addition to species-level predictions, other taxonomic levels could be generated by downgrading outputs to family or BMWP family level. The BMWP system is based on selected families of macroinvertebrates. Each family within the system is allocated a score in the range 1 to 10 according to its known tolerance to organic pollution, the most pollution-intolerant families being allotted the highest scores (Armitage, Moss *et al.* 1983). The BMWP score for a site is the sum of the scores of the BMWP families present in the sample, and the average score per taxon (ASPT) is simply the BMWP site score divided by the number of BMWP taxa. Expected values for BMWP score, number of BMWP taxa and ASPT were calculated from the expected probabilities of the BMWP families, and therefore observed/expected ratios were available for each BMWP index. These ratios were particularly suited to the requirements of water industry biologists engaged in routine biological monitoring. In 1986, the classification and prediction system was implemented on a simple microcomputer and made available to water industry biologists throughout Great Britain for testing under the acronym RIVPACS I. The software gave predictions at species, family and BMWP family levels and offered four reduced sets of predictive variables for appraisal. As a result of the testing exercise, the value of the prediction system was widely recognised and gave further impetus to the project.

Between 1984 and 1988, further sites on deep lowland rivers and on small streams (<5 km from source) were added, to give a total of 438 reference sites that were used to develop a new classification with 25 groups. The variables for prediction were reappraised and Water Authority/River Purification Board biologists requested single and paired season predictions. There was also progress in developing procedures for site evaluation based on observed to expected (O/E) ratios and the use of banding to distinguish unstressed sites (with O/E ratios close to unity) from stressed sites (with progressively lower O/E ratios). Note that the O/E ratio for each BMWP index is sometimes referred to as the Ecological Quality Index (EQI), as in Chapters 4 to 6 of this volume.

RIVPACS II

The newly formed National Rivers Authority (NRA) decided to fund development of the 438-site classification and prediction system into an operational version (RIVPACS II) for use in the 1990 River Quality Survey. RIVPACS II ran on IBM and IBM-compatible Personal Computers in addition to mainframe computers, and offered a number of significant improvements over the previous test version (Cox, Furse *et al.* 1991). It retained interactive predictions but, more importantly, it offered a datafile-operated system to make the prediction procedures more efficient. It also offered a menu of six sets of environmental variables for prediction (Wright *et al.* 1993), the choice of single, paired and three seasons combined predictions, plus a new variable taxonomic level (customisation) to provide compatibility with the requirements of local laboratories. Finally, it incorporated the improved method for the classification of new sites first suggested by Rushton (1987) (see Chapter 2).

RIVPACS II was used at almost 9000 sites in the 1990 River Quality Survey throughout England, Wales and Scotland, and on a more experimental basis in Northern Ireland, where there were no local reference sites (Sweeting, Lowson *et al.* 1992). The survey was conducted at BMWP family level and the initial output was as O/E ratios for the three BMWP indices. A banding system was then devised to distinguish good quality from progressively stressed sites in order to satisfy the reporting requirements of the national survey (Wright *et al.* 1993). Further discussion of the merits and difficulties inherent in devising a banding system may be found in Clarke, Furse *et al.* (1996).

This large-scale use of the system, and further detailed tests undertaken by the IFE, indicated that RIVPACS II performed effectively on many rivers. Nevertheless, there was a need to include additional river types, improve geographical coverage and increase the number of small stream sites to ensure that a future system could deliver reliable predictions throughout the country. The IFE tests also confirmed some inadequacies in prediction of the fauna (e.g in chalk streams) and demonstrated that a more rigorous protocol was required for accepting reference sites. Such problems were to be expected with a new methodology, but the potential advantages of this approach warranted the time and effort needed to refine the system (Royal Commission on Environmental Pollution 1992).

RIVPACS III

An upgraded version of the system was required for the 1995 General Quality Assessment (GQA) Survey, and data were collated for an additional 245 reference sites in Great Britain (Wright, Furse *et al.* 1995). These included sites of high biological quality previously sampled by the IFE, and further sites recommended by local biologists in order to make the dataset more comprehensive. The system was further enhanced by including 70 sites in Northern Ireland. As before, all species-level identifications were undertaken by IFE staff.

A large number of exploratory analyses were undertaken before RIVPACS III was finalised (Wright, Furse *et al.* 1995; Chapter 2). An early attempt to include all UK sites within a single classification was abandoned when it became apparent that the more restricted fauna in Northern Ireland would compromise both the Great Britain and Northern Ireland components of an all-inclusive system. Parallel classification and prediction systems were therefore developed for Great Britain and Northern Ireland within RIVPACS III. The enlarged dataset of 663 sites for Great Britain was progressively reduced, using several different procedures to ensure high quality in the final dataset. First, a series of criteria for site acceptance were applied to all sites. They included a minimum BMWP score of 100, a minimum of 30 standardised taxa and a nearest neighbour dissimilarity value not exceeding 0.55 (Belbin 1992). Next, during a series of classification and prediction exercises, sites that proved to be poor examples within their classification group (demonstrated by O/E ratios or BMWP indices) were rejected. Finally, five sites that had a surprisingly high taxon richness for their geographical location were removed because they were judged to be disruptive to a general prediction system (Wright, Furse *et al.* 1995). By this route, the dataset for Great Britain was reduced to 614 reference sites, of which just 386 of the original 438 RIVPACS II sites were retained.

Before developing RIVPACS III, Moss, Wright *et al.* (1999) demonstrated that several widely different approaches to site classification had potential as the starting point for a prediction system. However, when the 614-site Great Britain dataset became available, it was clear that a TWINSPAN classification would provide the best basis for the new prediction system (Wright, Furse *et al.* 1995; Chapter 2). Nevertheless, one important modification was implemented. Whereas RIVPACS II was based on qualitative species data, in RIVPACS III, qualitative species data plus family data (with $\log_{10}$ categories of abundance) were used to characterise each site (Wright, Furse *et al.* 1995; Wright, Moss *et al.* 1997). This new procedure, which used more information on the fauna, succeeded in creating more coherent site groupings of chalk stream sites, where problems had been encountered in RIVPACS II.

RIVPACS III+

The structure and major features of RIVPACS III are examined in the next section. However, one further development of the software, described by Clarke in Chapter 3, is listed in Table 1.1. RIVPACS III+ represents a major step forward through the incorporation of error terms for the O/E ratios used to assess site quality, and provides a mechanism for detecting statistically significant spatial and temporal differences between the macroinvertebrate assemblages of

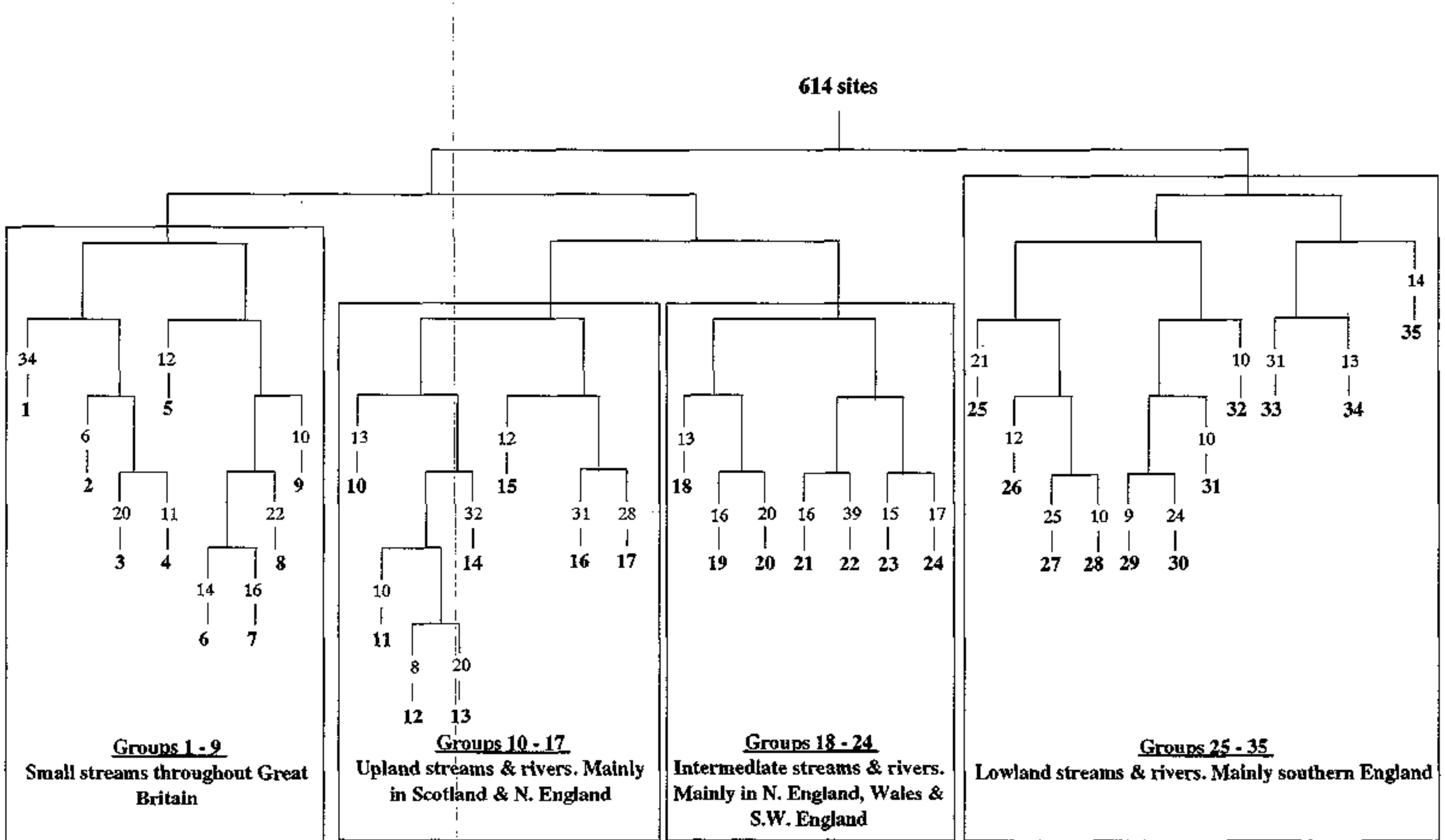


Figure 1.1. Classification of the 614 reference sites for Great Britain into 35 groups in RIVPACS III. The number of sites in each group is shown above each group number (the latter are given in **bold type**.) Four major blocks of sites are also indicated.

sites. In this way, it mirrors the objective procedures developed by the NRA for reporting chemical quality (National Rivers Authority 1994).

A review of RIVPACS III

Classification of sites

As previously indicated, two parallel classification and prediction systems were developed within RIVPACS III, based on 614 sites in Great Britain and a further 70 sites in Northern Ireland. In this review, attention will focus on the Great Britain component of the system. The 614 reference sites were divided into 35 classification groups, based on faunal composition (Fig. 1.1). Group size varied from six to 39 sites with a mean value of 17.5 sites per group, as in RIVPACS II. The steps leading to this end-result are presented elsewhere (Wright, Furse *et al.* 1995). Whereas RIVPACS II displayed a progression from upland to lowland groups, the RIVPACS III classification was more complex. Nevertheless, the 35 groups could be partitioned into four major blocks to provide a framework for interpreting the classification.

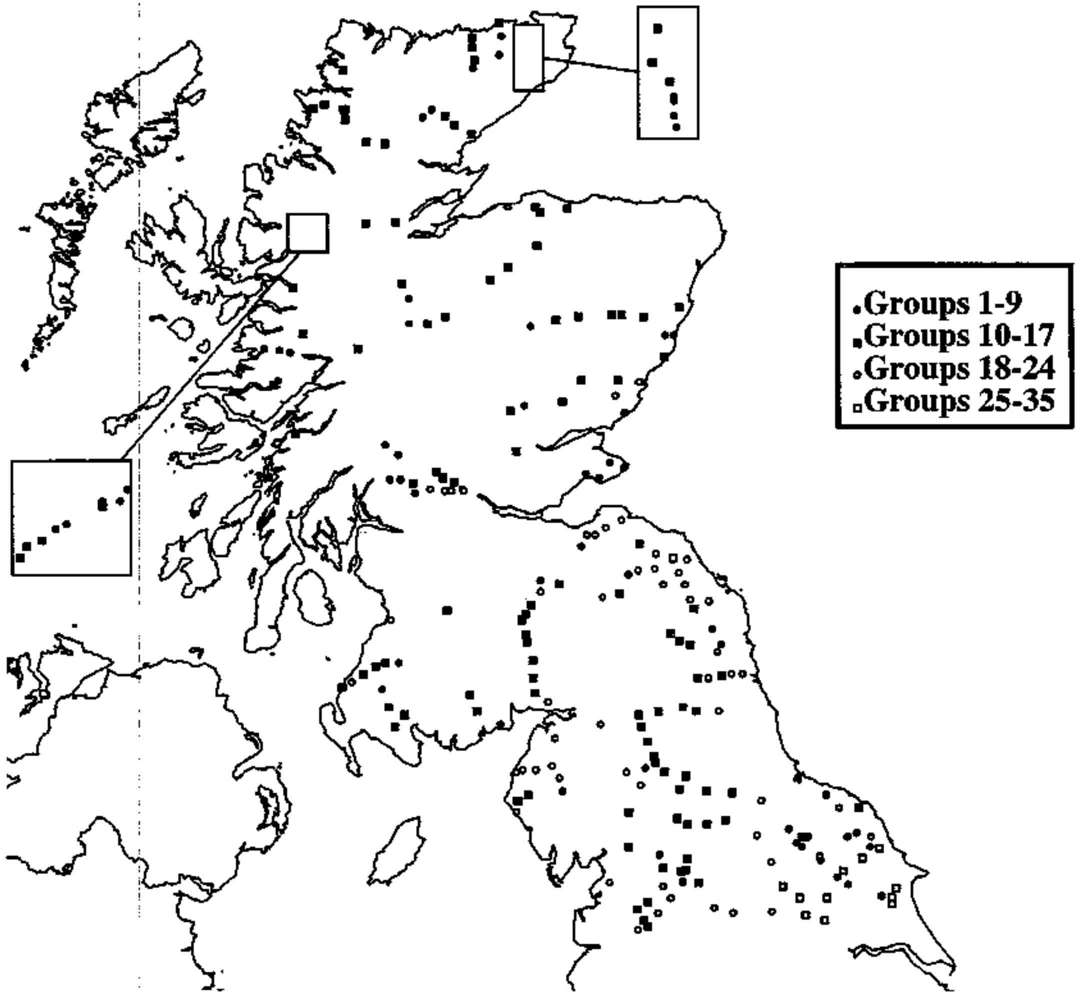
The 145 sites in Groups 1 to 9 were headwater and small stream sites which occurred throughout most of Great Britain, although central and south-eastern England were poorly represented. Groups 10 to 17, with 154 sites, were predominantly medium to large upland sites found mainly in Scotland and northern England. Groups 18 to 24 included 136 medium to large river sites in northern England, Wales and south-west England. Finally, Groups 25 to 35, with 179 sites, were lowland streams and rivers in the south and east. One feature of this large lowland block was the inclusion of three small stream groups (29, 31, 32) that occurred in the area where sites in Groups 1 to 9 were sparse. Thus, the additional small stream sites affected the overall structure of the earlier classifications. Nevertheless, the strong environmental contrasts between the north and west compared to the south and east were still apparent in the sequence of macroinvertebrate assemblages in Groups 10 to 17, 18 to 24 and 25 to 35 (Fig. 1.2).

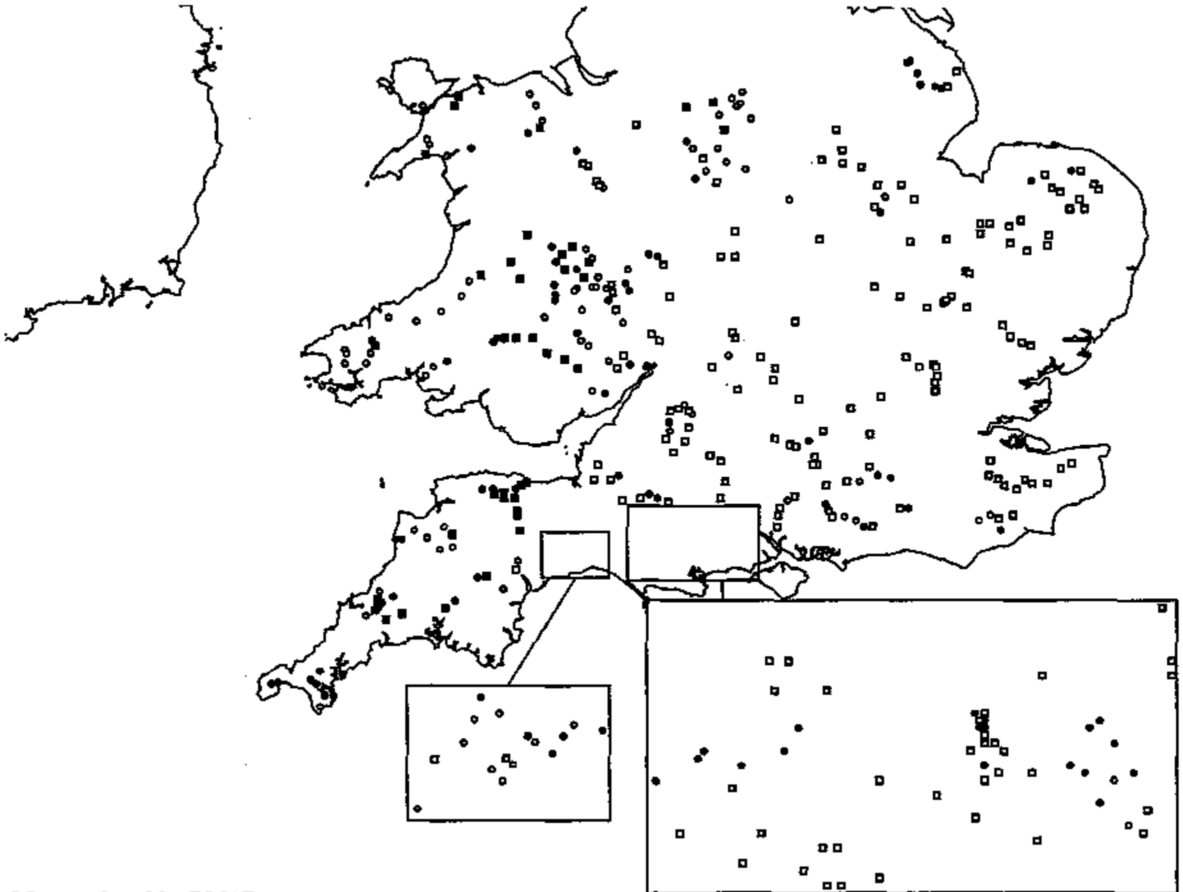
The mean standardised taxon richness for the 35 classification groups varied from 47 taxa in Group 5 to 109 taxa in Group 18 (Fig. 1.3). In general, richness was lower in the small stream groups (47–75 taxa per group) and in the upland groups (50–83 taxa) compared to the intermediate (72–108 taxa) and lowland groups (70–98 taxa). It was apparent that a wide variety of rivers in south Wales, and southern and eastern England, were capable of supporting taxon-rich assemblages. Species composition changed progressively across the classification but the frequency of occurrence of some non-insect groups was lower in the small stream (Groups 1–9) and upland (Groups 10–17) sections of the classification (Wright *et al.* 1998a).

At BMWP family level, the mean BMWP score per classification group varied from 255.2 in Group 18 to 123.5 in Group 5. The range for mean number of BMWP taxa varied from 40.1 taxa (Group 25) to 20.7 (Group 5), and ASPT ranged from 6.98 (Group 11) to 4.82 (Group 34).

System for predicting expected fauna at sampling sites

When the 614 sites in RIVPACS III were subjected to an internal MDA test using the standard environmental variables (Option 1 in RIVPACS III, Table 1.2), 51.6% of the sites were predicted to the "correct" group based on the re-substitution procedure. In this test, a correct prediction occurs when the classification group in which a site is placed with the highest probability of group membership, by MDA, is the same as the classification group in which the site was placed in the original TWINSPAN classification. Clearly, this is a severe test because there are 35 possible groups into which the site may be placed using the relevant environmental features in the MDA equations. If the classification is viewed as an artificial division of a


Groups 1-9
Groups 10-17
Groups 18-24
Groups 25-35



Map produced by DMAP

Figure 1.2. The location of the RIVPACS III classification Groups 1–9, 10–17, 18–24 and 25–35, in Great Britain.

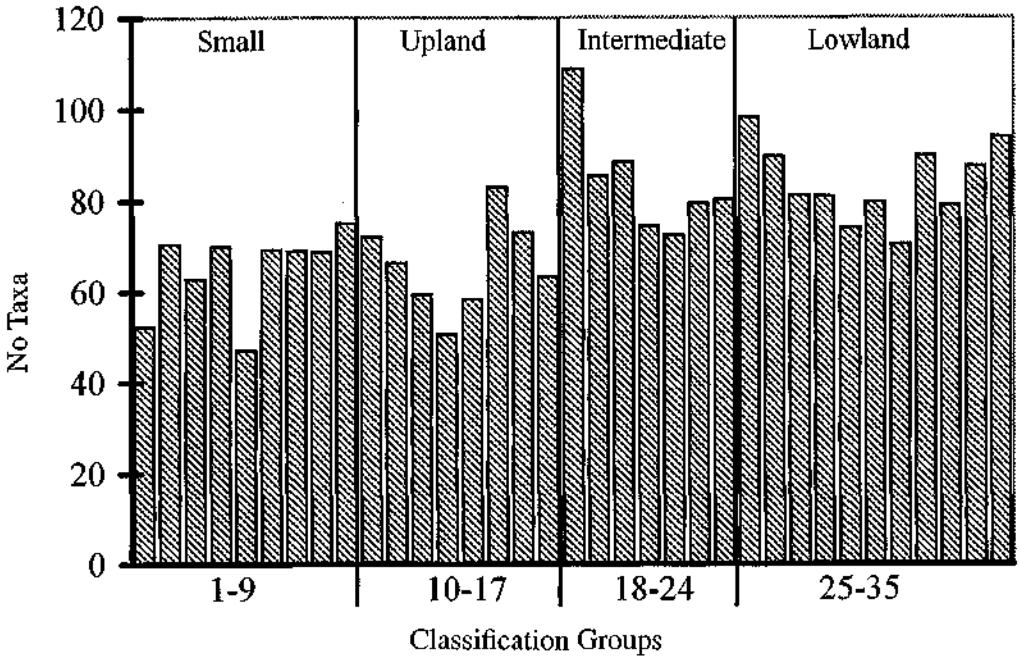


Figure 1.3. Mean standardised taxon richness for the 35 classification groups in RIVPACS III.

continuum of sites, then cases in which the second highest probability of group membership corresponds to the correct group should also be of value in the prediction system. A further 18.1% of sites were in this category.

Following a detailed assessment of O/E ratios and chi-squared tests to confirm that the correct taxa were being predicted (Wright, Furse *et al.* 1995), it was clear that RIVPACS III was setting higher standards for the expected fauna and BMWP index values than RIVPACS II, and performing at least as well as the previous system despite the fact that it now encompassed a wider range of river systems.

Table 1.2. *The five environmental options available for prediction in RIVPACS III. Option 1 is recommended for use in Great Britain.*

All five options require the following eight variables:

Distance from source (km)	Altitude (m)
Mean substratum (phi units)	Discharge category (9 categories, cumecs)
Mean water width (m)	Mean water depth (cm)
Latitude (°N)	Longitude (°W/°E)

Two to four additional variables are also required, according to the option chosen.

Variable	Option 1	Option 2	Option 3	Option 4	Option 5
Alkalinity (mg $CaCO_3 l^{-1}$)	+	+	–	+	–
Slope (m km^{-1})	+	–	+	+	–
Mean air temperature (°C)	+	+	+	–	+
Air temperature range (°C)	+	+	+	–	+

The suite of environmental variables for prediction and the taxonomic options available in RIVPACS III are closely similar to those in RIVPACS II (Wright *et al.* 1994) but in each case one or two changes have been made. The 12 environmental variables recommended for use in Great Britain (Option 1) remain unchanged (Table 1.2) but Option 6 from RIVPACS II has been deleted. The latter required the use of chloride as a predictor variable, but because high chloride could be an indicator of environmental stress, it was deemed inappropriate for use in RIVPACS.

The environmental data for a given site can be used to predict any or all of the taxonomic levels and seasonal options previously available in RIVPACS II (Table 1.3). One addition to RIVPACS III is the facility to test a new experimental index (Q14) (Wright, Furse *et al.* 1995), designed to detect early signs of environmental stress prior to major loss of taxon richness, as expressed through low O/E ratios. The index requires $\log_{10}$ abundance data for single seasons and operates on pollution-sensitive BMWP families with scores in the range 4 to 10. The Q14 index measures the observed loss of abundance for each family in relation to expectation, and a threshold value based on the reference dataset is used to indicate environmental stress.

Table 1.3. *The taxonomic, seasonal and prediction options available in RIVPACS III.*

(1) 3 = three seasons combined; 2 = any two seasons combined; 1 = spring, summer or autumn.

Taxonomic level	Seasonal options[(1)]	Type of prediction
Species	3, 2 or 1	Presence/absence only
All families	3, 2 or 1	Presence/absence only
Customised	3, 2 or 1	Presence/absence only
BMWP	3, 2 or 1	Presence/absence+biological indices
All	1 only	Log_{10} categories of abundance+Q14 index

In RIVPACS III, predictions can be made either interactively or in batch mode if the biological and environmental data are held in computer files. The observed fauna can then be compared with the expected fauna derived from RIVPACS using O/E ratios, and further categorisation of the results into bands can be made when required.

Examples of predictions based on RIVPACS III

This section includes three predictions to demonstrate the flexibility of the system. Each one is for a different location on the Moors River in Dorset, south-west England. Some sections of this river have been scheduled as a Grade 1 Site of Special Scientific Interest (SSSI) (Ratcliffe 1977) because they support rich assemblages of macrophytes and macroinvertebrates. However, other sections have been polluted by industrial and domestic effluents during a period of progressive urbanisation. An extensive survey of the macroinvertebrate fauna of this river in the mid-1980s (Wright, Welton *et al.* 1988) provides the data for these predictions.

1. A species-level prediction for an unstressed site

The Kings Farm site, in the middle reaches of the Moors River, is of high biological quality and forms one of the 614 reference sites in RIVPACS III. This species-level prediction demonstrates the procedure for comparing the observed with the expected fauna. Figure 1.4 lists the environmental data required for Option 1 of RIVPACS III in order to predict the probabilities of classification group membership. A site with these environmental features has the greatest affinity with Group 33, followed by Groups 32, 25 and 30. RIVPACS uses information on the probabilities of group membership, together with the frequency of

MOORS RIVER AT KINGS FARM

Mean width (m)	3.9	Altitude (m)	19
Mean depth (cm)	48.8	Distance from source (km)	17.0
Substratum (%)		Slope (m km^{-1})	1.4
Boulders + cobbles	0	Discharge category	3
Pebbles + gravel	9	Mean air temperature (°C)	10.57
Sand	16	Annual air temp. range (°C)	12.39
Silt and clay	74	Latitude (°N)	50.51
Mean substratum (phi)	6.01	Longitude (°W)	1.51
Alkalinity (mg l^{-1} $CaCO_3$)	161.7		

Classification groups predicted using the above data:-

Gp 33 = 72.1% ; Gp 32 = 17.8% ; Gp 25 = 4.3% ; Gp 30 = 3.8%.

Predicted taxa, in decreasing order of probability of capture:-

*	97.5%	Glossiphonia complanata (L.)
	97.4%	Micropsectra group
*	97.1%	Erpobdella octoculata (L.)
*	95.8%	Gammarus pulex (L.)
*	94.9%	Elmis aenea (Muller)
*	94.5%	Hydracarina
*	94.3%	Cricotopus group
*	92.6%	Pisidium nitidum Jenyns
*	90.4%	Pisidium subtruncatum Malm
*	89.8%	Psammoryctides barbatus (Grube)
*	88.6%	Asellus aquaticus (L.)
*	87.4%	Potamopyrgus jenkinsi (Smith)
	87.2%	Sphaerium corneum (L.)
*	86.1%	Aulodrilus pluriseta (Piguet)
	86.0%	Paratanytarsus group
*	85.4%	Thienemannimyia group
*	85.0%	Limnodrilus hoffmeisteri Claparede
*	83.1%	Helobdella stagnalis (L.)
	81.4%	Sialis lutaria (L.)
*	81.1%	Ceratopogonidae
*	78.5%	Polypedilum sp.
*	78.1%	Baetis vernus Curtis
*	78.1%	Lymnaea peregra (Muller)
*	77.7%	Physa fontinalis (L.)
	77.6%	Sigara (Sigara) sp.
*	75.5%	Stylaria lacustris (L.)
*	75.3%	Potamonectes depressus (Fabricius)
	75.3%	Procladius sp.
*	74.6%	Lumbriculus group
*	72.0%	Polycelis nigra group
*	72.0%	Microtendipes sp.
	70.6%	Gyraulus albus (Muller)
	70.2%	Centroptilum luteolum (Muller)
*	69.5%	Limnephilus lunatus Curtis
*	68.9%	Caenis luctuosa group
	66.1%	Anisus vortex (L.)
*	65.2%	Ephemerella ignita (Poda)
*	64.5%	Oulimnius tuberculatus (Muller)
*	64.4%	Hydroptila sp.
	63.3%	Bithynia tentaculata (L.)
*	62.4%	Eukiefferiella group
	61.7%	Valvata piscinalis (Muller)
*	59.9%	Ancylus fluviatilis Muller
*	59.8%	Crangonyx pseudogracilis Bousfield
*	57.8%	Piscicola geometra (L.)
	56.1%	Potamothrix hammoniensis (Michaelsen)
*	54.7%	Polycentropus flavomaculatus (Pictet)
*	52.0%	Halesus sp.
	51.9%	Sigara falleni (Fieber)
*	51.4%	Rhyacodrilus coccineus (Vejdovsky)
*	50.7%	Baetis rhodani (Pictet)
*	50.5%	Prodiamesa olivacea (Meigen

At 50% probability level: $\frac{\text{observed no.taxa}}{\text{expected no.taxa}} = \frac{39.0}{39.0} = 1.00.$

Figure 1.4. Species-level prediction for a site of high biological quality (Kings Farm, Moors River, Dorset, south-west England). Observed taxa are shown with an asterisk. See the text for further details.

occurrence of taxa in the relevant classification groups, to generate the taxa expected in the absence of environmental stress (Chapter 2). In practice, the list of expected taxa is presented in decreasing probability of capture and, if requested, can include all of the 637 taxa in the reference dataset for Great Britain. However, for species predictions, the list is frequently terminated at the 50% probability level (Fig. 1.4), to avoid printouts with large numbers of low probability taxa.

The taxa captured at Kings Farm by the standard sampling protocol (i.e. the observed fauna) are held in a separate biological file, but the RIVPACS printout identifies each observed taxon within the expected taxon listing using an asterisk. At sites of high quality, most taxa predicted with a high probability of capture should be observed at the site, but at the 50% probability level, only around one in two of the expected taxa should be observed after employing the standard protocol. To compare the observed fauna with the expected probabilities over the 100 to 50% probability range, the asterisks are summed to give the number of observed taxa (39), and all probabilities between 100 and 50% are summed to give the expected number of taxa (39.0) (see Fig. 1.4). At the 50% probability level, the O/E ratio for this site is unity, indicating that the site is of high biological quality.

2. BMWP family predictions at stressed sites

Chemical pollution and other forms of environmental stress affect individual species of macroinvertebrates. However, in Britain, the need for rapid appraisal of sites by non-specialists led the BMWP to develop a monitoring system based on selected families of macroinvertebrates (National Water Council 1981). As previously explained, expected values for the three BMWP indices can now be computed within RIVPACS for a given site with stated environmental features. When stress leads to the loss of some sensitive (higher scoring) families, then lowering of the O/E ratios for BMWP score, number of scoring taxa and ASPT will result.

A sampling site at Hurn, in the lower reaches of the Moors River, was subject to both organic and pesticide pollution in the mid-1980s. The environmental data for the site, the probabilities of group membership and the expected probabilities of family occurrences over the range 100 to 50% are presented in Fig. 1.5. Although most of the families expected with a very high level of probability (100 to 95%) are present, there is under-representation of all other families listed to 50% probability. In particular, several families of Trichoptera, Plecoptera and Ephemeroptera are missing. These detailed results are valuable to the local biologist when searching for evidence of the cause of the problem. However, it is also important to devise simple procedures for converting lists of taxa into a form which may be communicated to non-biologists and the public. This is achieved using BMWP indices.

In calculating BMWP family-level indices, the standard practice is to sum all the expected probabilities from 100% to 0.1%, and not simply for the 100 to 50% probability range illustrated in Fig. 1.5. For this reason, it is also critical to record the total number of families collected by the standard sampling protocol. At Hurn, four of these families had expected probabilities of occurrence below 50%. Figure 1.5 includes the O/E ratios for the three BMWP indices derived by RIVPACS from the observed and expected BMWP families. Whereas 24 BMWP taxa were observed, the expected number was 35.2, resulting in an O/E ratio well below unity (0.68). The O/E ratio for BMWP score was even lower (0.52), implying the loss of high scoring taxa, and this was confirmed by the O/E for ASPT which was also low (0.76). Although these results indicate an environmental problem, formal procedures are required to integrate the results from the O/E ratios and generate a quality band. These procedures are critical for reporting national surveys, targeting sites for remedial action, and for documenting changes over time. Some important considerations in the development of a scientifically credible banding system are presented by Clarke in Chapter 3, and a practical system devised for reporting the 1995 General Quality Assessment Survey is described by Hemsley-Flint in Chapter 4.

3. Abundance-based index (Q14)

So far, the emphasis has been on recognising the loss of individual species or BMWP families as a means of detecting stress. In practice, changes in the abundance of taxa also occur in response to stress, most notably when organic enrichment creates favourable conditions for a

MOORS RIVER AT HURN

Mean width (m)	11.3	Altitude (m)	10
Mean depth (cm)	30.3	Distance from source (km)	30.0
Substratum (%)		Slope (m km^{-1})	0.9
Boulders + cobbles	4	Discharge category	4
Pebbles + gravel	66	Mean air temperature (°C)	10.61
Sand	22	Annual air temp. range (°C)	12.32
Silt and clay	8	Latitude (°N)	50.46
Mean substratum (phi)	-1.38	Longitude (°W)	1.49
Alkalinity (mg l^{-1} $CaCO_3$)	92.7		

Classification groups predicted with the above data:-

Gp 19 = 57.7% ; Gp 30 = 9.4% ; Gp 28 = 7.1% ; Gp 25 = 6.3% ; Gp 32 = 4.7%
Gp 26 = 4.4% ; Gp 27 = 4.2%.

Predicted taxa, in decreasing order of probability of capture:-

*	100.0%	Chironomidae		78.3%	Hydroptilidae
*	100.0%	Oligochaeta		76.6%	Haliplidae
*	100.0%	Elmidae		74.5%	Leuctridae
*	100.0%	Baetidae	*	74.3%	Lymnaeidae
*	99.8%	Sphaeriidae		73.2%	Polycentropodidae
*	99.7%	Hydropsychidae		72.2%	Sialidae
*	99.7%	Simuliidae		71.5%	Leptophlebiidae
*	99.2%	Gammaridae	*	70.4%	Asellidae
*	97.4%	Hydrobiidae		70.0%	Sericostomatidae
*	97.2%	Tipulidae		65.9%	Gyrinidae
	96.8%	Ephemerellidae		61.6%	Goeridae
*	94.1%	Erpobdellidae	*	57.3%	Planariidae
	93.7%	Limnephilidae	*	57.3%	Planorbidae
*	90.4%	Leptoceridae		51.1%	Psychomyiidae
	89.7%	Caenidae	*	50.7%	Piscicolidae
*	87.3%	Glossiphoniidae	*	50.5%	Hydrophilidae
	86.5%	Dytiscidae			
	82.3%	Rhyacophilidae	[*	48.2%	Valvatidae
*	81.3%	Ancylidae	*	33.1%	Physidae
	80.4%	Heptageniidae	*	28.0%	Calopterygidae
	79.8%	Ephemeridae	*	18.9%	Taeniopterygidae]

Summary of BMWP Indices:-

	No.Taxa	BMWP Score	ASPT
Observed (O)	24	108	4.50
Expected (E)	35.2	208.9	5.93
O/E	0.68	0.52	0.76

Figure 1.5. BMWP family-level prediction for a stressed site (Hurn, Moors River, Dorset, south-west England), including the use of BMWP indices. See the text for further details.

restricted number of low-scoring BMWP families. In this case, many of the sensitive taxa have already disappeared and therefore O/E ratios based on presence/absence data are capable of exposing the problem without the need for quantitative data. There is a greater need for a procedure that can detect early effects of stress before loss of taxa leads to low O/E values derived from presence/absence data. RIVPACS III includes an experimental abundance-based index (Q14) that operates on single season family data with attached $\log_{10}$ categories of abundance. This index uses the proportional deficit of observed abundances below those expected for taxa with a BMWP score of 4 to 10, and ignores high abundance of BMWP taxa with a score of 1 to 3. In early tests it proved to be highly discriminatory amongst sites at the higher end of the quality spectrum but tended not to differentiate between sites with only low-scoring taxa. Ideally, it should be used at sites that are in a good quality band based on presence/absence data but where there are concerns over the early onset of environmental stress. The Q14 value below which stress may be indicated is based on the lower 5-percentile value of the index, using single season data for the 614 reference sites in RIVPACS III.

The Q14 index is demonstrated for a site below Palmersford Sewage Treatment Works, which is located between Kings Farm and Hurn. A preliminary combined season BMWP family-level prediction using presence/absence data indicated that the site just qualified as Biological Class A, based on criteria developed by the IFE and used in the 1990 River Quality Survey (Sweeting, Lowson *et al.* 1992). Figure 1.6 lists the observed and expected $\log_{10}$ abundance categories for the appropriate BMWP families in spring and includes the Q14 index, as determined by RIVPACS. The Q14 value of 32.43 is well below the critical limit of 42 and suggests that the site is stressed despite the previous Class A rating based on presence/absence data. However, more testing is required to fully determine the utility of this index.

The macroinvertebrate reference dataset

The biological data for the 614 reference sites included in RIVPACS III forms a unique historical record of the fauna at a wide range of running-water sites across Great Britain, because of the quality of the sites, the use of standard protocols in the field and laboratory, and the reliability of identifications. All the samples have been retained for future reference.

Taxon richness in the reference dataset

After combining the data for three seasons and applying the standard edits program, the standardised taxon richness varied between 31 and 134 taxa across the 614 reference sites. However, at over 80% of the sites, the number of standardised taxa per site fell within the range 50 to 99 (Wright *et al.* 1998a). Whereas extremely taxon-poor sites were confined to physically harsh environments in upland areas of Scotland and northern England, the most taxon-rich sites encompassed a wide range of river types in south Wales, southern England and East Anglia.

Wright *et al.* (1998b) present frequency distributions of the number of taxa per site at standardised "species", family and BMWP family levels. They also demonstrate that there were very highly significant correlations between the number of species and families ($r = 0.890$) and between the number of species and BMWP families ($r = 0.854$), as might be expected. During national and local surveys, a substantial number of sites in Great Britain are sampled at family or BMWP family level using the RIVPACS protocol. In view of these correlations, the data may contain valuable information for the early detection of species-rich sites of interest to the statutary nature conservation organisations within Great Britain.

The standardised taxon list for the 614 reference sites in Great Britain includes 637 "species", of which 142 are non-insects and 495 are insects (Table 1.4 and the Appendix). A

MOORS RIVER D/S PALMERSFORD STW

Mean width (m)	5.8	Altitude (m)	12
Mean depth (cm)	101.1	Distance from source (km)	25.0
Substratum (%)		Slope (m km^{-1})	1.0
Boulders + cobbles	0	Discharge category	4
Pebbles + gravel	2	Mean air temperature (°C)	10.61
Sand	27	Annual air temp. range (°C)	12.30
Silt and clay	72	Latitude (°N)	50.48
Mean substratum (phi)	6.17	Longitude (°W)	1.50
Alkalinity (mg l^{-1} $CaCO_3$)	98.5		

Classification groups predicted using the above data:-

Gp 33 = 82 % ; Gp 32 = 11 % ; Gp 25 = 2.4 % ; Gp 35 = 1.6 % ; Gp 30 = 1.5 %.

Comparison of expected (EXP) and observed (OBS) BMWP family log categories of abundance:-

EXP	OBS		EXP	OBS	
1.95	2	Gammaridae	.42	1	Coenagriidae
1.69	1	Baetidae	.42	0	Tipulidae
1.60	0	Elmidae	.37	0	Ephemeridae
1.54	0	Caenidae	.30	0	Psychomyiidae
1.45	0	Limnephilidae	.29	0	Calopterygidae
1.06	1	Leptoceridae	.29	0	Dendrocoelidae
1.01	2	Dytiscidae	.24	0	Sericostomatidae
.98	0	Simuliidae	.23	0	Goeridae
.96	2	Planariidae	.20	0	Rhyacophilidae
.83	0	Haliplidae	.20	0	Molannidae
.79	0	Corixidae	.19	2	Gyrinidae
.64	0	Polycentropodidae	.17	1	Piscicolidae
.60	1	Hydropsychidae	.16	0	Heptageniidae
.60	0	Hydroptilidae	.15	0	Neritidae
.60	0	Ancylidae	.15	0	Nemouridae
.55	0	Leptophlebiidae	.14	0	Hydrophilidae
.49	0	Ephemerellidae	.14	0	Lepidostomatidae
.46	0	Sialidae	.14	0	Notonectidae

Q14 = 32.43 (values below 42 may indicate stress).

Figure 1.6. BMWP family-level prediction using $\log_{10}$ category abundance data and the Q14 index for a site below Palmersford Sewage Treatment Works, Moors River Dorset, south-west England. See the text for further details.

full listing of the taxa, together with their frequency of occurrence in the 614 site dataset, is given by Wright, Blackburn *et al.* (1996). Apart from its value to river biologists as a source of information on the taxa most likely to be encountered in lotic waters, the list also flags threatened species with Red Data Book status (Shirt 1987; Bratton 1991) and rare species accorded "Nationally Scarce" status within Great Britain. Nationally Scarce taxa should be known from 100 or fewer 10x10 km squares of the National Grid (Bratton 1991). The RIVPACS III dataset includes 14 Red Data Book species (1 Gastropoda, 1 Bivalvia, 3 Ephemeroptera, 4 Coleoptera, 4 Trichoptera and 1 Diptera) and a further 47 which currently have the status of Nationally Scarce species (1 Gastropoda, 4 Bivalvia, 1 Ephemeroptera, 1 Plecoptera, 2 Odonata, 28 Coleoptera, 1 Megaloptera, 7 Trichoptera and 2 Diptera). (Note: In Wright, Blackburn *et al.* (1996), 15 RDB species were listed including two Gastropoda. A

Table 1.4. *Contribution of each major taxonomic group of freshwater invertebrates to the 637 standardised taxa recorded at the 614 RIVPACS III sites in Great Britain. Further details are given in the Appendix on page 24.*

(1) Taxonomic groups that are not identified to species.

(2) The level of identification varies from family to family.

Non-insects	No. of taxa	Insects	No. of taxa
Spongillidae(1)	1	Ephemeroptera	37
Hydridae(1)	1	Plecoptera	27
Tricladida	9	Odonata	13
Chordodidae(1)	1	Hemiptera	28
Ectoprocta(1)	1	Coleoptera	104
Gastropoda	29	Megaloptera	3
Bivalvia	22	Neuroptera	2
Aeolosomatidae(1)	1	Trichoptera	98
Oligochaeta	51	Lepidoptera(1)	1
Hirudinea	14	Diptera(2)	182
Hydracarina(1)	1		
Crustacea	11		
Totals:	142		495

careful re-examination of the single specimen of *Segmentina nitida* Müller (RDB1), which is now in poor condition, has failed to provide conclusive corroboration of the earlier identification and this must therefore be rejected). Over the past 20 years, the large scale of the sampling programme and the careful attention to species identifications has resulted in two Oligochaeta, three Ephemeroptera and one member of the Diptera being added to the British list.

There are further opportunities to utilise this dataset by examining the biological information in relation to environmental attributes. These could include the variables used in prediction, chemical determinands collected by the Environment Agency and the Scottish Environment Protection Agency, and further geological, geomorphological and other variables which may be accessed through a Geographic Information System.

Some future challenges

Refinement of RIVPACS is an on-going process, and each of five workshops held at Oxford (see Chapters 20 to 24) has focused on a topic which contributes to this process. In this concluding section, the emphasis is on items where progress is anticipated in the near future.

Early warning indices

Q14 is just one of fourteen indices devised by Ralph Clarke (see Wright, Furse *et al.* 1995) to provide early warning of major change in family abundance prior to substantial loss of families. Caution is required in the development and use of such indices because the RIVPACS sampling protocol is effort-dependent and takes account of all habitats, rather than involving quantitative sampling on one defined habitat. Nevertheless, some procedures which compare the observed and expected $\log_{10}$ abundance categories of families in a single season show promise, and further testing of abundance indices will attempt to identify those best able to detect the early stages of environmental stress.

New environmental features for prediction

The strong link between the environmental features of the reference sites and their biological assemblages has always been critical to the success of RIVPACS. Early in the project, care was taken to choose a limited set of environmental variables for prediction which were easy to acquire and had high predictive capability. Recently, the emphasis has been on the acquisition of a comprehensive set of reference sites that pass severe criteria for acceptance. It is now time to consider whether catchment characteristics, such as geological, soil, geomorphological and hydrological factors, can increase the reliability of the prediction system. It should be possible to obtain some attributes through a Geographic Information System. If the value of additional variables or alternatives to existing field-collected variables can be demonstrated, then a future version of RIVPACS may have broader application.

A link between RIVPACS and the River Habitat Survey

At present, RIVPACS is more successful at predicting species composition than the particular species richness to be expected at an individual site, when using the standard sampling protocol. In other words, the system generates the list of taxa with attached probabilities of occurrence, but local factors will influence the species richness observed at the site under study. At high quality sites the observed to expected ratio can be either above or below unity, and only when the ratio falls to a stated value below unity is the site regarded as stressed. In the RIVPACS classification, sites with similar species composition are grouped together, but within a given classification group the taxon richness varies from site to site. This is the result of genuine differences in the environmental conditions at the sites, the impact of natural stresses such as recent floods and droughts prior to sampling, and the fact that the sites will vary somewhat in their biological quality. The prediction methodology, in drawing on information from many sites across several classification groups, minimises the problems which would otherwise be encountered if the expected fauna was derived from a small number of local reference sites.

The addition of new predictive variables that describe local habitat features seems unlikely to offer a major breakthrough in the prediction of taxon richness in RIVPACS, which currently offers a general system for predicting the typical fauna to be expected across a very wide range of running-water sites under stated environmental conditions. An alternative approach might involve a detailed appraisal of the physical habitat of the 614 reference sites, to investigate variation in O/E ratios above and below unity. One possible route would be to link RIVPACS and the River Habitat Survey (RHS) approach (Raven, Fox *et al.* 1997) developed by the Environment Agency. In this method, the physical features of a series of semi-natural reference sites have been used to develop a classification (= typology) of rivers. This is the basis for predicting the semi-natural character of a river segment from map-derived data. Comparison of the observed physical character of the segment with the predicted semi-natural character provides an assessment of habitat quality. A comparison between habitat quality derived through the RHS and biological quality derived from RIVPACS could be instructive and offer a starting point for detailed investigations on site features that promote species richness.

Patterns in faunal composition

As previously indicated, the 614 reference sites represent a unique dataset and it is important to extract the maximum information, given the enormous effort required to assemble it. RIVPACS depends on the correlation of site attributes with faunal composition, for the purpose of prediction. The end-product is a system with considerable practical application, but it cannot explain the processes responsible for the observed patterns of faunal change along and between river types. In truth, this is a major challenge also requiring the use of experimental techniques. Nevertheless, the reference dataset provides an opportunity to investigate patterns in faunal composition among and within river types, in order to gain a wider perspective on the structure and functioning of minimally impacted sites.

Interpreting the cause of stress

Finally, although RIVPACS has well-developed methodologies for detecting stressed sites, interpretation of the cause of stress is still largely in the hands of the biologist with local knowledge and time to examine the differences between the list of observed and expected taxa. Clearly, there is potential for the development of new techniques for interpretation of results. These might include the development of new indices sensitive to specific forms of stress, an expert system for interpreting the differences between the observed and expected fauna, or the parallel use of ecotoxicological techniques to aid interpretation.

Acknowledgements

A number of organisations have provided financial support for the development of RIVPACS over the past 20 years. They include the Natural Environment Research Council, the Department of the Environment, the Scottish Office, the Welsh Office, the Department of the Environment (Northern Ireland), the National Rivers Authority and the Environment Agency. The RIVPACS team is grateful to each one of these organisations. The project also benefited from site data collected under contract to the former Nature Conservancy Council. Finally, it is a pleasure to acknowledge that RIVPACS is the end-product of a team effort by scientists at the River Laboratory of the Freshwater Biological Association and more recently the Institute of Freshwater Ecology, scientists from the Institute of Terrestrial Ecology, and biologists from the Environment Agency, the Scottish Environment Protection Agency and the Industrial Research & Technology Unit in Northern Ireland. I also thank an external reviewer and colleagues for helpful comments on the manuscript.

Appendix:– List of 22 major taxonomic groups of macroinvertebrates containing 117 families (and numbers of "standardised taxa"; total = 637) from the 614 reference sites for running-waters in Great Britain used in RIVPACS III (see Table 1.4 in this chapter; Wright, Blackburn *et al.* 1996; Wright *et al.* 1998a, 1999b).

Porifera
- Spongillidae (1)

Coelenterata
- Hydridae (1)

Tricladida
- Planariidae (5)
- Dugesiidae (2)
- Dendrocoelidae (2)

Nematomorpha
- Chordodidae (1)

Ectoprocta
- Ectoprocta (1)

Gastropoda
- Neritidae (1)
- Viviparidae (1)
- Valvatidae (3)
- Hydrobiidae (1)
- Bithyniidae (2)
- Lymnaeidae (5)
- Physidae (3)
- Planorbidae (11)
- Acroloxidae (1)
- Ancylidae (1)

Bivalvia
- Margaritiferidae (1)
- Unionidae (2)
- Sphaeriidae (18)
- Dreissenidae (1)

Aphanoneura
- Aeolosomatidae (1)

Oligochaeta
- Naididae (20)
- Tubificidae (22)
- Enchytraeidae (2)
- Haplotaxidae (1)
- Lumbriculidae (5)
- Lumbricidae (1)

Hirudinea
- Piscicolidae (1)
- Glossiphoniidae (7)
- Hirudinidae (1)
- Erpobdellidae (5)

Hydracarina
- Hydracarina (1)

Crustacea
- Argulidae (1)
- Asellidae (2)
- Corophiidae (1)
- Crangonyctidae (1)
- Gammaridae (4)
- Niphargidae (1)
- Astacidae (1)

Ephemeroptera
- Siphlonuridae (2)
- Baetidae (13)
- Heptageniidae (5)
- Leptophlebiidae (6)
- Ephemerellidae (2)
- Potamanthidae (1)
- Ephemeridae (3)
- Caenidae (5)

Plecoptera
- Taeniopterygidae (3)
- Nemouridae (9)
- Leuctridae (6)
- Capniidae (2)
- Perlodidae (3)
- Perlidae (2)
- Chloroperlidae (2)

Odonata
- Platycnemididae (1)
- Coenagrionidae (5)
- Calopterygidae (2)
- Gomphidae (1)
- Cordulegastridae (1)
- Aeshnidae (2)
- Libellulidae (1)

Hempitera
- Mesoveliidae (1)
- Hydrometridae (1)
- Veliidae (1)
- Gerridae (1)
- Nepidae (1)
- Naucoridae (1)
- Aphelocheiridae (1)
- Notonectidae (3)
- Corixidae (18)

Coleoptera
- Haliplidae (11)
- Noteridae (1)
- Dytiscidae (36)
- Gyrinidae (6)
- Hydraenidae (11)
- Hydrophilidae (22)
- Scirtidae (4)
- Dryopidae (2)
- Elmidae (11)

Megaloptera
- Sialidae (3)

Neuroptera
- Osmylidae (1)
- Sisyridae (1)

Trichoptera
- Rhyacophilidae (4)
- Glossosomatidae (2)
- Philopotamidae (3)
- Polycentropodidae (9)
- Ecnomidae (1)
- Psychomyiidae (6)
- Hydropsychidae (9)
- Hydroptilidae (6)
- Phryganeidae (2)
- Limnephilidae (19)
- Molannidae (1)
- Beraeidae (3)
- Odontoceridae (1)
- Leptoceridae (23)
- Goeridae (3)
- Lepidostomatidae (3)
- Brachycentridae (1)
- Sericostomatidae (2)

Lepidoptera
- Pyralidae (1)

Diptera
- Tipulidae (31)
- Psychodidae (19)
- Ptychopteridae (1)
- Dixidae (5)
- Chaoboridae (1)
- Culicidae (1)
- Thaumaleidae (1)
- Ceratopogonidae (1)
- Chironomidae (86)
- Simuliidae (19)
- Stratiomyidae (3)
- Empididae (4)
- Dolichopodidae (1)
- Rhagionidae (3)
- Tabanidae (2)
- Syrphidae (1)
- Ephydridae (1)
- Sciomyzidae (1)
- Muscidae (1)